

AMENDMENTS TO THE SPECIFICATION**In the specification:**

Please substitute the following paragraph for the originally filed paragraph appearing in the specification on page 1, line 5, under the heading "RELATED APPLICATION:"

The instant application is a continuation under 37 CFR 1.53(b) of pending U.S. Application No. 09/654,365, filed September 1, 2000, which is hereby incorporated herein by reference in its entirety. The benefit of an earlier filing date under 35 U.S.C. § 120 is claimed.

Please substitute the following paragraph for the originally filed paragraph appearing in the specification on page 6, lines 1-17:

FIG. 1 is a diagram of a partially expanded view of an exemplary computer environment 100 in which the features and aspects of the present invention may be implemented. Computer environment 100 includes memory 102, central processing unit (CPU) 104, input device 106, I/O controller 108, video display 110, and secondary storage device 112. Memory 102 contains IR system 114. Secondary storage device 112 contains training documents 116 and testing documents 122. Documents may include articles (e.g., stories) from a newswire, radio/television audio broadcast (speech recognition engine needed), articles from a digital library, web sites on the world wide web, or any other files or data that are identifiable by their association with one or more topics. A topic is one or more words or phrases specifying an area of interest. For example, with respect to a news story, a topic could be defined as a particular event, such as specific bombings, elections, crimes, trials, etc. A user may specify a topic through a

query to IR system 114. The user may be operating a remote computer (not shown) and may send the query to computer environment 100 via I/O controller 108. The remote computer may be connected to computer environment 100 via the Internet, modem dialup, on-line service, ISDN, wireless communication, or other data transmission scheme. Alternatively, the user may be operating computer environment 100 locally, using, for example, input device 106.

Please substitute the following paragraph for the originally filed paragraph appearing in the specification on page 9, lines 10-20:

A simple example of the data flow depicted in FIG. 2 follows. Suppose that there are 500 training documents, and that 4 of those documents have been labeled as being related to a particular trial. The training documents can be input to training module 204, where a model for the trial is generated using at least the 4 documents related to the trial. This model along with all 500 training documents are input to training document track score module 208, where a raw score 214 is generated for each of the training documents. Of all of the raw scores 214, 496 of the scores relate to off-topic documents, and 4 of the scores relate to on-topic documents. The model is also input to testing document track score module 212 along with a testing document 210, for which a decision as to whether or not the document relates to the trial is desired. As a result a raw score 220 is generated for the testing document 210. For the purposes of this example, assume that the raw score 220 is 8.5.

Please substitute the following paragraph for the originally filed paragraph appearing in the specification on page 14, lines 4-19:

Notice that the normalized score is based on statistics relating to off-topic document scores. Previous efforts at score normalization tend to focus on on-topic documents. Typically, however, there are many more off-topic documents for a given topic than there are on-topic documents. Another way of thinking about on-topic documents is that they are “not off-topic.” By basing the score normalization on statistics relating to off-topic documents, a more accurate decision on whether a testing document is on-topic or off-topic can be made, because there is more statistical data available for what constitutes an off-topic document than is available for what constitutes an on-topic document. Moreover, on-topic documents were used to build the model, so score normalization based on only on-topic documents would inherently be biased. Generally, a low normalized score indicates that the document is not much different from the training documents that were designated as off-topic documents and thus should be designated as off-topic. A high normalized score, however, suggests that the document is more likely to be different from the off-topic training documents and thus should be designated as on-topic. A low or high normalized score, however, does not necessarily guarantee that a testing document will be judged as off-topic or on-topic, respectfully. Other factors weigh in, and the final determination depends on a normalized score’s comparison to a threshold.